Optimising Roadwork Project Timing - Statement of Work (SOW)

DATA201/422

2024-09-20

Overview

Fulton Hogan conduct significant roadworks in busy city areas across the country. This can be disruptive to city businesses, costly due to traffic management, and dangerous for workers in high traffic areas. Fulton Hogan is therefore engaged with us, Bad Ass Analytics Limited (Baa Ltd), to investigate if school holidays may be a more appropriate period to plan roadworks in city areas. The particular cities of interest are Auckland, Wellington, and Christchurch. However, they wish to be handed a complete and cleaned dataset so their own Data Scientists can conduct further investigations in future.

Key questions

- Does it make sense to plan roadworks for school holidays in CBD areas ?
- Which days are best if roadworks must be completed during the day ?
- Are there any geographical differences between CBDs we must be aware of ?

Data

Baa Ltd has access to telecommunications data from Spark and Vodafone in the lead up and during school holidays in 2024. These will be used to determine the change in population densities in the three relevant CBDs. The data in its entirety, once cleaned, will be provided to the client as detailed in the deliverables.

Supporting data has been made available alongside the telecommunications data, and contains everything required to complete the project. You may not need all the data provided, but you will not require any additional data. Here is a brief summary of datasets provided:

- `sp_data.csv.gz` - Telecommunications data from Spark
- `vf_data.parquet` - Telecommunications data from Vodafone
- `sa2_2023.csv` - SA2 codes and names
- `sa2_ta_concord_2023.csv` - Concordance for SA2 to Territorial Authorities (TA)
- `subnational_pop_ests.csv` - A file with population estimates with demographic breakdown
- `urban_rural_to_indicator_2023.csv` - Concordance between urban/rural codes and their type
- `urban_rural_to_sa2_concord_2023.csv` - Concordance between urban/rural and SA2 codes

Deliverables

Code repository

The code repository must be a version controlled repository that can be easily shared with the client. In this case GitHub, with the group name in the repository name, under the DATA422 GitHub organisation. It must be completed in the R programming language, used by Fulton Hogan analysts.

A clean dataset

The dataset must be outputted as a single gzipped CSV. It will contain the Territorial Authority code, the Statistical Area Level 2 code, a datetime in NZST, and a count of people in the area.

Readme on said repository

This is a Markdown document that is rendered on the GitHub repository page describing the code. It will outline the purpose, and list each script with a brief description on what it is doing and any requirements. All datasets being used will also be described so that the codebase can be maintained by the clients Data Scientists.

A PDF report detailing methods

The methods document is for a technical audience. It should not be too long but describes the entire pipeline in detail. Any corrections, conversions, or clarifications around the data should be included. Charts could be useful where there are specific details you wish to highlight. Also include a section on limitations of the data, and a section on future improvements to consider for future development. Once again, this is to be handed over to the clients Data Scientists so this is who you need to keep in mind.

A PDF report addressing key questions from stakeholders

This is for a non-technical audience within the executive leadership team at Fulton Hogan. This information may also be shared with other stakeholders, such as councils, to justify certain timing, or support the consenting process. The report will address the key questions mentioned in this SOW, and will include any supporting charts to aide a non-technical reader in their decision making. **You are not expected to conduct any statistical tests to address key questions.**

Trello board with screenshots for setup and weekly progress

Project management must be conducted in order to see a thorough and timeline project. Proving the use of a project management tool is obviously less realistic in a real scenario but oh well, sue me. Once the DATA422 students have set up an initial project, each card must be assigned to a person. At the end of each week, you will screen shot the board to show how the project is tracking along (there is no marks based on number of cards or what is moving). If there are any roadblocks or updates for each card (such as something could not be completed, is blocked, or encountered a challenge). Then there must be a comment updating the cards throughout the project. You must submit the setup, then weekly screenshots of the project, contained in the GitHub repository. Include a link to the repository in your Readme and we will check any relevant comments and assignment of tasks.

Risks and Assumptions

Students not pulling their weight

The most significant risk is a slack student. However, we are mitigating that risk by adjusting grades based on Git commit history, assignment of Trello cards, and any comments on those Trello cards.

Issues with group dynamic

If a student is at risk of breaching the student code of conduct, then report to a lecturer (Sinéad) immediately. However, anything short of this we would encourage students to try manage issues between themselves. Part of this project is giving a 'real world' experience. Managing the group dynamic is potentially a positive experience you can take through to job interviews and later work. However, we will reiterate, that this does not extend to students who breach the code of conduct (this includes harassment, bullying, discrimination, etc.). If you suspect a breach, then a lecturer can help determine whether further action is required.